AD-A104 479 JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL--ETC F/G 12/1
STRONG CONVERGENCE OF NEAREST NEIGHBOR REGRESSION ESTIMATORS.(U)
JUN 81 P E CHENG N00014-79-C-0801
UNCLASSIFIED TR-347 NL

1 OF 1
AD A
104479

END
DATE
FILMED
10-81
DTIC

AD A104479

9

LEVEL II

DEPARTMENT OF MATHEMATICAL SCIENCES
The Johns Hopkins University
Baltimore, Maryland 21218

STRONG CONVERGENCE OF NEAREST NEIGHBOR REGRESSION ESTIMATORS[1]

by

Philip E. Cheng
University of Houston

Technical Report No. 347
ONR Technical Report No. 81-7
June ~~1981~~ 1981

DTIC
ELECTE
SEP 23 1981
S D
D

Research supported by the Army, Navy and Air Force under Office of Naval Research Contract N00014-79-C-0801. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[1] Research performed at The Johns Hopkins University.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

81 9 22 028

DTIC FILE COPY

ABSTRACT

STRONG CONVERGENCE OF NEAREST NEIGHBOR REGRESSION ESTIMATORS

Let X be R^d-valued and Y be real valued in the framework of nonparametric estimation of a regression function $R(x) = E(Y|X = x)$. The uniform measure of deviation $\| T_n - R\|_B = \sup_{x \in B} |T_n(x) - R(x)|$ is studied for estimators T_n of the nearest neighbor type. It is shown that $\| T_n - R\|_B \to 0$ almost surely if the conditional variance of Y given X, $Var(Y|X)$, is a bounded random variable. The associated rate of convergence $\| T_n - R\|_B = o(n^{(\delta-1)/(2+d)})$, any $\delta > 0$, is obtained assuming that $E|Y|^{2+d} < \infty$, $Var(Y|X)$ is a bounded random variable, and R is Lipschitz of order 1.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

1. Introduction. Consider estimation of the regression function $R(x) = E(Y|X = x)$ given a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ from an unknown joint distribution function F. The idea of using nonparametric estimates of the nearest neighbor type was initiated by Fix and Hodges (1951) in their study of nonparametric discriminatory analysis. The following definition of *nearest neighbor regression estimator*, considered by Royall (1966), Stone (1977) and Devroye (1978), is adopted in this note. Let X be R^d-valued and Y be real-valued. For each x in R^d, order the (X_i,Y_i), $i = 1,\ldots,n$, according to nondecreasing distances $\| X_i - x\|$ (here, $\|\cdot\|$ denotes the usual Euclidean distance on R^d) and obtain $(X^x_{1n},Y^x_{1n}),\ldots,(X^x_{nn},Y^x_{nn})$ where X^x_{1n} is the nearest X_i to x and X^x_{nn} the farthest. In case of ties of distances, the order may be arbitrarily determined among the subsets of ties. Define an estimator of $R(x)$ by

$$T_n(x) = \sum_{i=1}^{n} c_{in} Y^x_{in}, \tag{1.1}$$

where the weights c_{in}, $i = 1,\ldots,n$ *are selected to satisfy*

Condition 1

(i) $\sum_{i=1}^{n} c_{in} = 1$, $c_{1n} \geq c_{2n} \geq \cdots \geq c_{nn} \geq 0$;

(ii) $\sum_{i=k(n)+1}^{n} c_{in} \to 0$ for a nondecreasing sequence $k(n)$ satisfying

$k(n) \to \infty$ and $k(n)/n \to 0$,

(iii) $\max_{1 \leq i \leq k(n)} c_{in} \to 0$.

In particular, the weights of the so-called *K-nearest neighbor* (K-NN) estimator satisfy

Condition 2

(i) $\sum_{i=1}^{n} c_{in} = 1$, $c_{1n} \geq \dots \geq c_{nn} \geq 0$,

(ii) $c_{in} = 0$ for $i > k(n)$, $k(n) \to \infty$ and $k(n)/n \to 0$,

(iii) $c_1/k(n) \leq c_{in} \leq c_2/k(n)$ for $i = 1,\dots,k(n)$ and positive constants

$0 < c_1 \leq 1 \leq c_2$.

Royall (1966) studied the pointwise mean square error of the estimate T_n. Stone (1977) characterized the weight functions $\{c_{in}, i = 1,\dots,n\}$ for which a form of mean L_p-consistency is achieved. Devroye (1978) established a uniform strong convergence result. Discussion of extensive further work related to nearest neighbor rules appears in Stone (1977, Section 9).

In general, the results available for estimators satisfying Condition 2 also hold for those satisfying Condition 1 provided that the quantity $\sum_{i=k(n)+1}^{n} c_{in}$ is sufficiently small. Examples of the K-NN weights (see Stone 1977) include the uniform weights with $c_{in} = 1/k(n)$, $i = 1,\dots,k(n)$, the triangular weights, and the quadratic weights. An interesting simplified version of (1.1) given by Devroye (1978) is the estimator

(1.2) $\quad \hat{T}_n(x) = T_n(X_{1n}^x)$.

Besides facilitating computation, the estimator $\hat{T}_n$ possesses the same strong consistency properties as T_n.

In this note, study is concentrated on the uniform almost sure convergence of the estimators T_n and $\hat{T}_n$. Under a standard condition (see Stone 1980) that the conditional variance of Y given X, $Var(Y|X)$, is a bounded random variable, it is shown that $\|T_n - R\|_B \to 0$ and $\|\hat{T}_n - R\|_B \to 0$ with probability 1, where B is assumed to be the bounded support of X in R^d. For any $\delta > 0$, the associated rates of convergence, $\|T_n - R\|_B = o(n^{(\delta-1)/(2+d)})$

and $\|\hat{T}_n - R\|_B = o(n^{(\delta-1)/(2+d)})$ with probability 1, are obtained under the additional regularity conditions that R is Lipschitz of order 1 and X possesses a density function f which is bounded away from zero on B.

2. <u>Results</u>. Let $(X_1,Y_1),\dots,(X_n,Y_n)$ be a random sample from a joint distribution function. Let P_X be the probability measure on R^d corresponding to the marginal distribution function of X. For each x in R^d, let $S^x(r) = \{u: \| u - x \| < r\}$, where $\|\cdot\|$ denotes the usual Euclidean metric. Define the support of X to be the set

$$B = \{x \in R^d: P_X(S^x(r)) > 0 \text{ for every } r > 0\}.$$

Then B is a closed subset of $(R^d, \|\cdot\|)$ (Devroye (1978, Lemma 1). For each $x \in B$, define $G^x(r) = P_X(S^x(r))$ and $R_{in}^x = \| X_{in}^x - x\|$. Thus $G^x(R_{in}^x)$, $i = 1,\dots,n$, are distributed as U_{in}, $i = 1,\dots,n$, the order statistics of a random sample of size n from the uniform distribution over the interval [0,1]. Following Devroye (1978), we say that the noise is in $L_t (t > 0)$ if there exists a finite positive constant D_t such that $\sup_{x \in B} E(|Y - R(x)|^t | X = x) \leq D_t$. We shall assume that the support B is a bounded (hence compact) subset of R^d. For simplicity, "k(n)" will be replaced with "k" denoting the greatest integral part of k(n). The symbol "c" denotes a generic positive constant which need not be the same at each appearance. The statement "with probability 1" is abbreviated as "w.p.1". Also, we shall omit the superscript "x", writing S(Y), G(Y), X_{in} and R_{in}. The following regularity conditions will be imposed.

<u>Assumption 1</u>

(i) R is continuous on B.

(ii) $E|Y|^t < \infty$,

(iii) The noise is in L_t.

We now state the main result for the estimator T_n.

THEOREM 1. *Suppose that Assumption* 1 *holds with* $t = 2$ *and that Condition* 2 *holds with* $k = n^{1/2}\beta_n \log n$, *where* $\beta_n \to \infty$ *arbitrarily*. *Then*

$$\| T_n - R\|_B \to 0 \quad \underline{w.p.}1.$$

Regarding the simplified version $\hat{T}_n$, we have

COROLLARY 1. *Under the conditions of Theorem* 1,

$$\| \hat{T}_n - R\|_B \to 0 \quad \underline{w.p.}1.$$

Devroye (1978) showed the strong consistency of the estimators T_n and $\hat{T}_n$ assuming that the noise is in L_t for $t > d+3$ and $t > 2$, respectively. A natural question raised by him is whether the results are valid if the noise is in L_2. The above theorem provides an affirmative answer.

Our next result discusses the associated rate of the uniform convergence in Theorem 1. We assume that X has a density function f. Choose any fixed positive number λ, and let $B_n = \{x \in B\colon \| x - y\| \geq \lambda n^{-1/(2+d)}$ for all y on the boundary of B$\}$. Thus, each B_n is a compact subset of B and B_n approximates B as $n \to \infty$. The following mild regularity conditions are imposed.

Assumption 2

(i) R is Lipschitz of order 1 on B, i.e. $|R(x) - R(y)| \leq c\| x - y\|$ for all x, y in B, for some positive constant c.

(ii) $\inf_{x\in B} f(x) \geq \mu$ for a positive constant μ.

(iii) $E|Y|^{2+d} < \infty$.

(iv) The noise is in L_2.

THEOREM 2. *Let Assumption* 2 *and Condition* 2 *hold with* $k = n^{2/(2+d)}$. *Then with probability* 1

(2.1) $$\| T_n - R\|_{B_n} = o(n^{(\delta-1)/(2+d)}) \quad \textit{for any } \delta > 0.$$

For the simplified version $\hat{T}_n$, we have

COROLLARY 2. *Under the conditions of Theorem* 2,

$$\| \hat{T}_n - R\|_{B_n} = o(n^{(\delta-1)/(2+d)}), \quad \textit{with probability } 1.$$

From Theorem 3 of Devroye (1978) it can be derived that (2.1) holds under a stringent L_t-noise condition with $t \geq 2 + (2+d)^2$. Also, from Theorem 4 of Devroye (1978) it can be checked that for $k = c(n \log n)^{1/2}$,

$$P(\| \hat{T}_n - R\|_B > \epsilon) \leq cnk^{-1}e^{-cn\epsilon^{2d}} + cnk^{1-t}.$$

Then $\| \hat{T}_n - R\|_B = O(\log n/n)^{1/2 d}$ with probability 1 under the L_t-noise condition for $t > 5$. This rate is faster than that in (2.1) only for $d \leq 2$ and at the expense of a more restrictive L_t-noise condition. Discussion of even more further restrictive *exponential type of noise conditions* is given in Devroye (1978).

3. Proofs. Without loss of generality, it suffices to prove the theorems for the case of uniform K-NN weights, i.e. $c_{in} = 1/k$, $1 \leq i \leq k$. Theorem 1 will be proved utilizing the following lemmas.

Consider for each n a permutation function $\sigma(i, x_1, \ldots, x_n)$, $i = 1, \ldots, n$, which for each given $\underline{x} = (x_1, \ldots, x_n)$ maps $\{1, \ldots, n\}$ onto $\{1, \ldots, n\}$. We have the following lemma (Royall (1966, Lemma 2.1)) on independence and conditionality.

LEMMA 1. *If* $(X_1,Y_1), \ldots, (X_n,Y_n)$ *are independent*, *then*

$$P\{Y_{\sigma(i)} \in B_i,\ i = 1,\ldots,n \mid X_{\sigma(i)} = x_{\sigma(i)},\ i = 1,\ldots,n\}$$
$$= \prod_{i=1}^{n} P\{Y_{\sigma(i)} \in B_i \mid X_{\sigma(i)} = x_{\sigma(i)}\} \quad \underline{w.p.}1.$$

In view of the definition (1.1) of T_n, Lemma 1 justifies use of the following terminology. We will let

$$V_n(x) = E(T_n(x) \mid X_{in},\ i = 1,\ldots,n)$$
$$= \frac{1}{k}\sum_{i=1}^{k} E(Y_{in} \mid X_{in}) = \frac{1}{k}\sum_{i=1}^{k} R(X_{in}) \quad w.p.1.$$

LEMMA 2. (Devroye (1978, Theorem 1)) *Assume Condition* 1 *and* (i) *of Assumption* 1. *Then*

$$\sup_{x\in B} R^x_{k,n} \to 0 \text{ and } \| V_n - R\|_B \to 0 \text{ w.p.1.}$$

In order to utilize a truncation argument, we define

$$\bar{T}_n(x) = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{in},$$

where

$$\bar{Y}_{in} = Y_{in}\, I[|Y_{in}| \le n^{1/2}],\ 1 \le i \le n.$$

Set

$$\bar{V}_n(x) = E(\bar{T}_n(x)|X_{in},\ i = 1,\ldots,n) = \frac{1}{k}\sum_{i=1}^{k} \bar{R}(X_{in}),$$

where

$$\bar{R}(X_{in}) = E(\bar{Y}_{in}|X_{in}),\ 1 \le i \le k.$$

LEMMA 3. *Assume* (ii) *of Assumption* 1 *with* $t = 2$. *Then*

$$\| T_n - \bar{T}_n\|_B \to 0, \text{ w.p.1.}$$

PROOF. (ii) of Assumption 1 implies that $\sum_{j=1}^{\infty} P(Y_j^2 > j) \le EY_1^2 < \infty$, from which it follows that

$$P(|Y_j| > j^{1/2} \text{ i.o.}) = P(Y_j^2 > j \text{ i.o.}) = 0.$$

Consequently, there exists a full set Ω such that for each $\omega \in \Omega$ there exists a finite positive integer N_ω such that $Y_j(\omega) > j^{1/2}$ only if $1 \le j \le N_\omega$. Set $J_\omega = \{Y_j(\omega),\ 1 \le j \le N_\omega\}$ for each $\omega \in \Omega$. Then $Y_{in}(\omega) \ne \bar{Y}_{in}(\omega)$ only if $Y_{in}(\omega) \in J_\omega$. Since J_ω is a finite set for each $\omega \in \Omega$ independent of $x \in B$, it follows that for each $\omega \in \Omega$

$$\| T_n(x,\omega) - \bar{T}_n(x,\omega)\|_B \le \frac{1}{k}\sum_{i=1}^{k} \| Y_{in}(\omega) - \bar{Y}_{in}(\omega)\|_B$$

$$\le \frac{1}{k}\sum_{Y_{in}(\omega)\in J_\omega} |Y_{in}(\omega) - \bar{Y}_{in}(\omega)| \to 0. \qquad \square$$

LEMMA 4. *Suppose* (i) *and* (iii) *of Assumption* 1 *hold with* $t = 2$. *Then*

$$\| V_n - \bar{V}_n\|_B \to 0 \text{ w.p.1.}$$

PROOF. From continuity of *R and boundedness of* B, $\| R\|_B < \infty$. For $t = 2$, (iii) of Assumption 1 implies that with probability 1

$$\| V_n - V\|_B \le \frac{1}{k}\sum_{i=1}^{k} \| R(X_{in}) - \bar{R}(X_{in})\|_B$$

$$= \frac{1}{k}\sum_{i=1}^{k} \| E(Y_{in} I(|Y_{in}| > n^{1/2})|X_{in})\|_B$$

$$\le \frac{1}{k}\sum_{i=1}^{k} \| E(Y_{in}^2|X_{in})\|_B\, n^{-1/2}$$

$$\le n^{-1/2}(D_2 + \| R\|_B^2) \to 0. \qquad \square$$

LEMMA 5. *Under the assumptions of Theorem* 1 *we have*

$$\| \bar{T}_n - \bar{V}_n\|_B \to 0 \text{ w.p.1.}$$

For the proof of Lemma 5, the following fact about independent variables *is useful.*

LEMMA 6. (Lamperti (1966)) *Let* $X_1,\ldots,X_n$ *be independent random variables with* $|X_i| \le M$, $EX_i = 0$ *and* $\mathrm{Var}(X_i) \le \sigma^2$ *for* $i = 1,\ldots,n$. *Let* $S_n = \sum_{i=1}^{n} X_i$. *Then for each* t, $0 \le t \le 2/M$,

$$E(\exp tS_n) \le \exp[nt^2\sigma^2(1+tM)/2].$$

PROOF OF LEMMA 5. It is seen from Devroye (1978, Theorem 3) that, through a combinatorial argument of Cover (1965) and Vapnik and Chervonenkis (1971) (see also Devroye and Wagner (1977)), the following *inequality holds:*

$$P\{\sup_{x\in B}(\bar{T}_n(x) - \bar{V}_n(x)) > \epsilon\}$$

$$\le c_d n^{d+1} \sup_{(x_1,\dots,x_k)\in B^k} P\{\frac{1}{k}\sum_{i=1}^{k}(\bar{Y}_i - \bar{R}(X_i)) > \epsilon \mid X_i = x_i,\ i = 1,\dots,k\}$$

$$\le c_d n^{d+1-\epsilon\beta_n} \sup_{x\in B^k} E\{\exp[\sum_{i=1}^{k}\frac{\beta_n \log n}{k}(\bar{Y}_i - \bar{R}(X_i))] \mid X_i = x_i,\ i = 1,\dots,k\}$$

$$\le c_d n^{d+1-\epsilon\beta_n} \sup_{x\in B^k} \prod_{i=1}^{k} E\{\exp[\frac{\beta_n \log n}{k}(\bar{Y}_i - \bar{R}(X_i))] \mid X_i = x_i\}$$

for a constant C_d depending only on the dimension d. By Lemma 1, $\{\bar{Y}_i - \bar{R}(X_i)\}_{i=1}^n$ given $\{X_i\}_{i=1}^n$ are conditionally independent with mean zero and finite variance, for $\sup_{x\in B} \mathrm{Var}(\bar{Y}_i|X_i) \le \sup_{x\in B} \mathrm{Var}(Y_i|X_i) \le D_2$. Therefore, Lemma 6 is applicable. For each factor in the product of (3.1) $|\bar{Y}_i - \bar{R}(X_i)| \le 2|\bar{Y}_i| \le 2n^{1/2}$ holds with probability 1 independently of $x \in B$. Set $t = \beta_n \log n/k$ and $M = 2n^{1/2}$ such that $t \le 2/M$. The right-hand side of (3.1) is bounded above by $c_d n^{d+1-\epsilon\beta_n} \prod_{i=1}^{k} \exp(3D_2 t^2/2)$ which is then bounded by $cn^{d+1-\epsilon\beta_n}$ asymptotically since $kt^2 \to 0$ as $n \to \infty$. Likewise, one can obtain

$$P\{\inf_{x\in B}(\bar{T}_n(x) - \bar{V}_n(x)) < -\epsilon\} \le cn^{d+1-\epsilon\beta_n}.$$

As $\beta_n \to \infty$ Lemma 5 follows via the Borel-Cantelli lemma. □

PROOF OF THEOREM 1. From the triangle inequality

$$\| T_n - R\|_B \le \| T_n - V_n\|_B + \| V_n - R\|_B$$

$$\le \| T_n - \bar{T}_n\|_B + \| \bar{T}_n - \bar{V}_n\|_B + \| \bar{V}_n - V_n\|_B + \| V_n - R\|_B$$

$$\to 0 \text{ with probability 1,}$$

according to Lemmas 2, 3, 4 and 5. □

The proof of Corollary 1 is omitted since it is a matter of defining analogous terminology as needed in the proof of Theorem 1 and checking straightforwardly that Lemmas 2, 3, 4 and 5 have valid counterparts when the estimator T_n is replaced by $\hat{T}_n$. We remark that the L_2-noise condition *is needed in the proof of Lemma 5 through an application of Lemma 6.* However, it is not known whether Theorem 1 (or Corollary 1) remains valid under any other weaker conditions, say the L_t-noise condition for $t = 1$.

We now turn to the proof of Theorem 2, for which it suffices to establish the following statements. For any $\delta > 0$, we have with probability 1

(3.2) $\quad n^{(1-\delta)/(2+d)} \| V_n - R\|_{B_n} \to 0,$

(3.3) $\quad n^{1/(2+d)} \| T_n - \bar{T}_n\|_B \to 0,$

(3.4) $\quad n^{(1-\delta)/(2+d)} \| V_n - \bar{V}_n\|_B \to 0,$ and

(3.5) $\quad n^{1/(2+d)} (\beta_n \log n)^{-1} \| T_n - \bar{V}_n\|_B \to 0.$

The idea is simply to provide rates of convergence for the results of Lemmas 2, 3, 4 and 5. Here, an alteration is noted: the analogous definitions of $\bar{T}_n$ and $\bar{V}_n$ are given by $\bar{T}_n = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{in}$ with $\bar{Y}_{in} = Y_{in} I[|Y_{in}| \le n^{1/(2+d)}]$, $i = 1,\dots,n$; and $\bar{V}_n = \frac{1}{k}\sum_{i=1}^{k} \bar{R}(X_{in})$ with $\bar{R}(X_{in}) = E(\bar{Y}_{in}|X_{in})$.

To show (3.2), we first observe that, by (i) of Assumption 2,

$$\| V_n - R\|_{B_n} \le \frac{1}{k}\sum_{i=1}^{k} \| R(X_{in}) - R(x)\|_{B_n} \le C\| R_{kn}\|_{B_n}.$$

Thus (3.2) follows if it can be shown that

(3.6) $\quad n^{(1-\delta)/(2+d)} \| R_{kn}\|_{B_n} \to 0$ with probability 1.

To show (3.6), let M be a bound on the diameter of the set B. For any $\epsilon > 0$,

let $v_n = \varepsilon n^{(\delta-1)/(2+d)}$. Thus $S^x(v_n) \subset B$ and $G^x(v_n) \geq c(d)\mu v_n^d$ for each $x \in B_n$, as soon as n is sufficiently large, where c(d) is the volume of the d-dimensional unit sphere. For each n, let $N_n = (M/2)^d/(v_n/2)^d \leq cv_n^d$. There exists a subset $\{x_j, 1 \leq j \leq N_n\}$ of B_n such that for each x in B_n there is at least one x_j with $\| x - x_j \| < v_n/2$. Thus $R_{kn}^x \geq v_n$ implies that $R_{kn}^{x_j} \geq v_n/2$. Consequently, for each positive integer m we have

$$\sum_{n=1}^{\infty} P[n^{(1-\delta)/(2+d)} \| R_{kn} \|_{B_n} > \varepsilon]$$

$$\leq \sum_{n=1}^{\infty} P[c\| R_{kn}\|_{B_n} > v_n]$$

$$\leq \sum_{n=1}^{\infty} P[\bigcup_{j=1}^{N_n} (cR_{kn}^{x_j} > v_n/2)]$$

$$\leq \sum_{n=1}^{\infty} P[\bigcup_{j=1}^{N_n} (G(R_{kn}^{x_j}) \geq G(cv_n))]$$

$$\leq \sum_{n=1}^{\infty} N_n P(G(R_{kn}) \geq cv_n^d)$$

$$\leq \sum_{n=1}^{\infty} N_n E(U_{kn}^m)/cv_n^{dm}$$

$$\leq \sum_{n=1}^{\infty} cn^{d(1-\delta)/(2+d)+m[2/(2+d)-1+d(1-\delta)/(2+d)]}$$

$$< \infty$$

provided that m is large enough that $m\delta > (2-\delta) + (2/d)$. Since m is arbitrary, this establishes (3.6) via the Borel-Cantelli lemma. Hence, (3.2) is proved.

For (3.3) we observe from the proof of Lemma 3 that the analogous term J_ω (now given by the fact that $E|Y|^{2+d} < \infty$) is a finite set for each ω in a full set Ω. Hence, for each $\omega \in \Omega$,

$$n^{1/(2+d)} \| T_n(x,\omega) - \bar{T}_n(x,\omega) \|_B$$

$$\leq n^{1/(2+d)} k^{-1} \sum_{i=1}^{k} |Y_{in}(\omega) - \bar{Y}_{in}(\omega)|$$

$$\leq n^{1/(2+d)} k^{-1} \sum_{Y_{in}(\omega) \in J_\omega} |Y_{in}(\omega) - \bar{Y}_{in}(\omega)|$$

$$\leq n^{-1/(2+d)} \sum_{Y_{in}(\omega) \in J_\omega} |Y_{in}(\omega) - \bar{Y}_{in}(\omega)| \to 0.$$

Thus (3.3) is proved.

For (3.4), following the argument of Lemma 4, it is seen that with probability 1

$$n^{(1-\delta)/(2+d)} \| V_n - \bar{V}_n \|_B$$

$$\leq n^{(1-\delta)/(2+d)} k^{-1} \sum_{i=1}^{k} \| E(Y_{in} I(|Y_{in}| > n^{1/(2+d)}) | X_{in}) \|_B$$

$$\leq n^{(1-\delta)/(2+d)} k^{-1} \sum_{i=1}^{k} \| E(Y_{in}^2 | X_{in}) \|_B n^{-1/(2+d)}$$

$$\leq n^{-\delta/(2+d)} (D_2 + \| R \|_B^2) \to 0.$$

Finally, to obtain (3.5), it suffices to show that for any $\alpha > 0$

$$\sum_{n=1}^{\infty} P\{n^{1/(2+d)} (\beta_n \log n)^{-1} \| \bar{T}_n - \bar{V}_n \|_B > \alpha\} < \infty. \tag{3.7}$$

From the proof of Lemma 6, it is checked that

$$P\{n^{1/(2+d)} (\beta_n \log n)^{-1} \sup_{x \in B} (\bar{T}_n(x) - \bar{V}_n(x)) > \alpha\} \tag{3.8}$$

$$\leq c_d n^{d+1-\alpha\beta_n} \sup_{\underset{\sim}{x} \in B^k} \prod_{i=1}^{k} E\{\exp\{n^{1/(2+d)} k^{-1} (\bar{Y}_{in} - \bar{R}(X_i))\} | X_i\}.$$

Now, for each factor of the product on the right-hand side of (3.8) it is seen that $|\tilde{Y}_i - \tilde{R}(X_i)| \le 2|\tilde{Y}_i| \le 2n^{1/(2+d)}$ w.p.1. Let $M = 2n^{1/(2+d)}$ and $t = n^{1/(2+d)}k^{-1} \le 2/M$. Lemma 6 is applicable since the noise is in L_2. Thus (3.8) is bounded above by $c_d n^{d+1-\alpha\beta_n} \prod_{i=1}^{k} \exp(3D_2 t^2/2)$ and hence bounded by $cn^{d+1-\alpha\beta_n}$ since $kt^2 \le 1$. Likewise,

$$P(n^{1/(2+d)}(\beta_n \log n)^{-1} \inf_{x \in B} (\tilde{T}_n(x) - \tilde{V}_n(x)) < -\alpha) \le cn^{d+1-\alpha\beta_n}.$$

Thus (3.7) holds as $\beta_n \to \infty$. (3.5) is proved.

PROOF OF THEOREM 2. Combining the results of (3.2), (3.3), (3.4) and (3.5), Theorem 2 is proved. $\square$

An analogous proof for Corollary 2 is also omitted.

ACKNOWLEDGEMENT. The author expresses his sincere thanks to Professor R. J. Serfling for several helpful comments.

REFERENCES

Breiman, L. (1968). Probability. Addison-Wesley.

Cover, T. M. (1965). Geometric and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. Electronic Computers, Vol. EC-14, 326-334.

Devroye, L. P. (1978). The uniform convergence of nearest neighbor regression function estimates and their application in optimization. IEEE Trans. Info. Th., IT-24, No. 2, 142-151.

Devroye, L. P. and Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. Ann. Statist., 5, No. 3, 536-540.

Fix, E. and Hodges, J. L. Jr. (1951). "Discriminatory analysis, nonparametric discrimination: consistency properties". Report No. 4, Project No. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Lamperti, J. (1966). Probability. W. A. Benjamin, Inc., N.Y.

Royall, R. M. (1966). A class of nonparametric estimators of a smooth regression function. Ph.D. dissertation, Stanford Univ.

Stone, C. J. (1977). Consistent nonparametric regression. Ann. Statist., 5, 595-645.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist., 8, 1348-1360.

Vapnik, V. N. and Chervonenkis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. Theory of Prob. and its Appl., Vol. 16, 264-280.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER ONR No. 347	2. GOVT ACCESSION NO.
3. RECIPIENT CATALOG NUMBER	
4. TITLE STRONG CONVERGENCE OF NEAREST NEIGHBOR REGRESSION ESTIMATORS	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORGANIZATION REPORT NO. Technical Report No. 347
7. AUTHOR(s) Philip E. Cheng	8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-79-C-0801
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematical Sciences The Johns Hopkins University Baltimore, Maryland 21218	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217	12. REPORT DATE June, 1981
	13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Nonparametric regression; nearest neighbor estimator; strong convergence.

20. ABSTRACT

Let X be R^d valued and Y be real valued in the framework of nonparametric estimation of a regression function $R(x) = E(Y|X = x)$. The uniform measure of deviation $\| T_n - R\|_B = \sup_{x\in B} |T_n(x) - R(x)|$ is studied for estimators T_n of the nearest neighbor type. It is shown that $\| T_n - R\|_B \to 0$ almost surely if the conditional variance of Y given X, $Var(Y|X)$, is a bounded random variable. The associated rate of convergence $\| T_n - R\|_B = o(n^{(\delta-1)/(2+d)})$, any $\delta > 0$, is obtained assuming that $E|Y|^{2+d} < \infty$, $Var(Y|X)$ is a bounded random variable, and R is Lipschitz of order 1.

DA
FIL

0